

SISTEM REKOMENDASI FILM MENGGUNAKAN ITEM-BASED CLUSTERING HYBRID METHOD

SKRIPSI

KONSENTRASI KOMPUTASI CERDAS DAN VISUALISASI

Diajukan untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer



Disusun Oleh :

RACHMAD TSALAATSA

NIM. 0610960053

PROGRAM STUDI INFORMATIKA / ILMU KOMPUTER

PROGRAM TEKNOLOGI INFORMASI DAN ILMU KOMPUTER

UNIVERSITAS BRAWIJAYA

MALANG

2013

SISTEM REKOMENDASI FILM MENGGUNAKAN ITEM-BASED CLUSTERING HYBRID METHOD

SKRIPSI

KONSENTRASI KOMPUTASI CERDAS DAN VISUALISASI

Diajukan untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer



Disusun Oleh :

RACHMAD TSALAATSA

NIM. 0610960053

**PROGRAM STUDI INFORMATIKA / ILMU KOMPUTER
PROGRAM TEKNOLOGI INFORMASI DAN ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA**

MALANG

2013

i

LEMBAR PERSETUJUAN

**SISTEM REKOMENDASI FILM MENGGUNAKAN
ITEM-BASED CLUSTERING HYBRID METHOD**

SKRIPSI

KONSENTRASI KOMPUTASI CERDAS DAN VISUALISASI

Diajukan untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer



Disusun Oleh :

RACHMAD TSALAATSA

NIM. 0610960053

Telah diperiksa dan disetujui oleh Dosen Pembimbing

Dosen Pembimbing I

Dosen Pembimbing II

Edy Santoso, S.Si, M.Kom
NIP. 19740414 200312 1 004

Dian Eka Ratnawati, S.Si, M.Kom
NIP. 19730619 200212 2 001

LEMBAR PENGESAHAN

**SISTEM REKOMENDASI FILM MENGGUNAKAN
ITEM-BASED CLUSTERING HYBRID METHOD**

SKRIPSI

KONSENTRASI KOMPUTASI CERDAS DAN VISUALISASI

Diajukan untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer

Disusun Oleh :

RACHMAD TSALAATSA

NIM. 0610960053

Skrripsi ini telah diuji dan dinyatakan lulus pada tanggal 17 Mei 2013

Penguji I

Penguji II

Lailil Muflikhah, S.Kom, M.Sc
NIP. 19741113 200501 2 001

Suprpto, ST., MT
NIP. 19710727 199603 1 001

Penguji III

Novanto Yudistira, S.Kom., M.Sc
NIP. 831110 16 1 1 0425

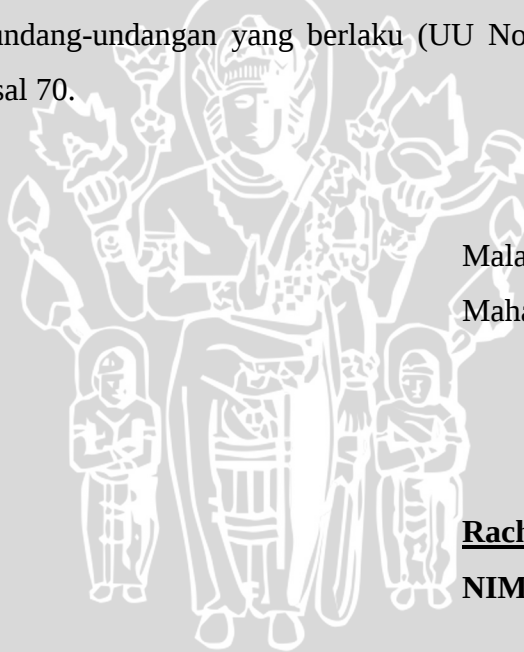
Mengetahui,
Ketua Program Studi Informatika/Ilmu Komputer

Drs. Marji, M.T
NIP. 19670801 199203 1 001

LEMBAR PERNYATAAN

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah SKRIPSI ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah SKRIPSI ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia SKRIPSI ini digugurkan dan gelar akademik yang telah saya peroleh (SARJANA) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).



Malang, 2 Mei 2013

Mahasiswa,

Rachmad Tsalaatsa

NIM. 0610960053

KATA PENGANTAR

Alhamdulillah, segala puji dan syukur penulis ucapkan kepada Allah SWT karena atas rahmat dan hidayah-Nya sehingga Penulis dapat menyelesaikan skripsi ini sebagai salah satu persyaratan untuk menyelesaikan studi di Program Studi Ilmu Komputer Program Teknologi Informasi dan Ilmu Komputer (PTIIK) Universitas Brawijaya.

Penulis menyadari bahwa skripsi ini dapat terselesaikan atas bantuan, petunjuk, dan bimbingan dari berbagai pihak yang telah membantu proses penyelesaiannya. Oleh karena itu, penulis mengucapkan terima kasih banyak kepada :

1. Edy Santoso, S.Si, M.Kom selaku Dosen pembimbing I yang telah memberikan bimbingan dan arahan dengan penuh kesabaran sehingga skripsi ini dapat terselesaikan.
2. Dian Eka Ratnawati, S.Si, M.Kom selaku Dosen pembimbing II yang telah memberikan bimbingan dan arahan dengan penuh kesabaran sehingga skripsi ini dapat terselesaikan.
3. Drs. Marji, MT selaku Ketua Program Studi Informatika/Ilmu Komputer Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya.
4. Dewi Yanti Liliana S. Kom, M.Kom selaku Dosen Penasehat Akademik.
5. Ir. Sutrisno, MT, selaku Ketua Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya.
6. Segenap Bapak dan Ibu Dosen yang telah mendidik dan mengajarkan ilmunya kepada Penulis selama menempuh pendidikan di Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya.
7. Segenap staf dan karyawan di Program Teknik Informatika Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya yang telah banyak membantu Penulis dalam pelaksanaan penyusunan skripsi ini.
8. Orang tua dan kakak-kakak yang tak henti-hentinya memberikan dukungan dan doanya sehingga terselesaikannya skripsi ini.

9. Seluruh rekan kerja PPTI Universitas Brawijaya yang telah membimbing dan memberi saran kepada penulis hingga terselesaikannya skripsi ini.
10. Teman-teman Ilmu Komputer angkatan 2006 tercinta yang telah banyak memberikan bantuan dan pengalaman selama menjadi mahasiswa di Universitas Brawijaya.
11. Semua pihak yang telah membantu terselesaikannya skripsi ini yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa skripsi ini masih banyak kekurangan dan jauh dari sempurna, karena keterbatasan materi dan pengetahuan yang dimiliki penulis. Maka, saran dan kritik yang membangun dari semua pihak sangat diharapkan demi penyempurnaan selanjutnya. Semoga skripsi ini dapat bermanfaat dan berguna bagi semua pihak, baik penulis maupun pembaca, dan semoga Allah SWT meridhoi dan dicatat sebagai ibadah. Amin.

Malang, 2 Mei 2013

Penulis



ABSTRAK

Sistem rekomendasi merupakan suatu sistem yang bertujuan memberikan saran kepada user tentang item yang akan mereka pilih. Pada industri film terdapat 2.482.179 judul film internasional dan jumlah tersebut akan terus bertambah. Dengan jumlah film yang banyak, hal itu membuat user menghadapi kesulitan dalam mencari dan memilih film mana yang tepat untuk dilihat sehingga dibutuhkan sistem rekomendasi.

Item-based Clustering Hybrid Method (ICHM) adalah salah satu metode dalam sistem rekomendasi yang menggabungkan metode collaborative filtering dan content-based. Dengan menggunakan ICHM, sistem dapat merekomendasikan item yang belum pernah di rating sama sekali oleh user.

Data film yang digunakan berasal dari GroupLens. Analisis tingkat akurasi prediksi rating dilakukan dengan menggunakan Means Absolute Error (MAE). Dari hasil analisis didapatkan penambahan jumlah cluster pada proses clustering tidak mempengaruhi tingkat akurasi hasil prediksi rating dengan nilai MAE sebesar 0,74. Dan penambahan nilai koefisien c pada proses similarity memberikan hasil prediksi yang semakin tidak akurat dengan nilai MAE tertinggi sebesar 0,744, sehingga dikatakan prediksi rating yang dihasilkan semakin menjauhi rating sebenarnya atau hasil prediksi rating semakin tidak akurat seiring bertambahnya koefisien c .

Kata Kunci : Sistem Rekomendasi, *ICHM*, Film, *MAE*

ABSTRACT

Recommendation system is a system aimed to give advice to user regarding the items that will be chosen. In movie industry, there are 2.482.179 titles of international movie and the number will keep on growing. The increasing number of movies will cause user to find difficulties in choosing the right movie to watch; thus, a recommendation system is needed.

Item-based Clustering Hybrid Method (ICHM) is a method in recommendation system combining collaborative filtering and content-based method. Using ICHM, system can recommend the items that have never been rated by user.

The movie data that is used comes from GroupLens. Analysis of rating prediction accuracy is performed using Means Absolute Error (MAE). From the analysis result, it is found that addition of the amount of cluster on clustering process doesn't affect the level of accuracy of rating prediction with MAE value of 0,74. And addition of c coefficient value on similarity process give an increasingly inaccurate prediction result with highest MAE value of 0,744, so that it is said that the resulting rating prediction is getting farther from the real rating, or the resulting rating prediction will be more inaccurate along with the increase of c coefficient.

Keywords : recommendation system, ICHM, movie, MAE

DAFTAR ISI

	Halaman
SAMPUL	i
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN	iv
KATA PENGANTAR	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Batasan Masalah	3
1.5 Manfaat Penelitian	3
1.6 Metodologi Penelitian	3
1.7 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	
2.1 Sistem Rekomendasi	5
2.2 <i>Content-based</i>	6
2.3 <i>Collaborative Filtering</i>	7
2.4 Metode <i>Hybrid</i>	8
2.5 <i>Item-based Clustering Hybrid Method</i>	9

2.5.1 Pembentukan Matriks.....	10
2.5.2 Perhitungan Nilai <i>Similarity</i>	12
2.5.2 Perhitungan Prediksi <i>Item</i>	13
2.6 Pengukuran Tingkat Akurasi.....	14

BAB III METODE PENELITIAN DAN PERANCANGAN

3.1 Studi Literatur	16
3.2 Data Penelitian	16
3.3 Deskripsi Umum Sistem	17
3.4 Perancangan Proses.....	17
3.4.1 Pembentukan Matriks Rating.....	19
3.4.1.1 Matriks <i>Item-rating</i>	20
3.4.1.2 Matriks <i>Group-rating</i>	21
3.4.2 Perhitungan Nilai <i>Similarity</i>	26
3.4.2.1 <i>Similarity Item-rating</i>	26
3.4.2.2 <i>Similarity Group-rating</i>	29
3.4.2.3 <i>Linier Combination of Similarity</i>	31
3.4.3 Perhitungan Prediksi Rating.....	32
3.4.3.1 Prediksi Pada <i>Item</i> yang Pernah di Rating	32
3.4.3.2 Prediksi Pada <i>Item</i> yang Belum Pernah di Rating	35
3.4.3.3 Hasil Prediksi	37
3.4.3.4 Rekomendasi <i>Non-Personalized</i>	37
3.4.3.5 Rekomendasi <i>Personalized</i>	37
3.5 Perancangan Antarmuka.....	37
3.5.1 Antarmuka Sistem.....	38
3.5.2 Antarmuka <i>Home</i>	38
3.5.3 Antarmuka Detail <i>Item</i>	39
3.5.4 Antarmuka Rating <i>User</i>	40
3.5.5 Antarmuka Rekomendasi.....	41
3.6 Perancangan Tabel	42
3.6.1 Tabel Data	42
3.6.2 Tabel <i>Item</i>	43

3.6.3 Tabel <i>User</i>	44
3.7 Perancangan Uji Coba.....	44
3.7.1 Bahan Pengujian.....	44
3.7.2 Tujuan Pengujian	45
3.7.3 Skenario Pengujian.....	45
3.8 Perhitungan Manual	47

BAB IV IMPLEMENTASI

4.1 Lingkungan Implementasi.....	62
4.1.1 Lingkungan Perangkat Keras	62
4.1.2 Lingkungan Perangkat Lunak	62
4.2 Implementasi Program	63
4.2.1 Implementasi <i>Clustering</i>	63
4.2.2 Implementasi <i>Similarity Item-rating</i>	66
4.2.3 Implementasi <i>Similarity Group-rating</i>	68
4.2.4 Implementasi <i>Linier Combination of Similarity</i>	69
4.2.5 Implementasi Prediksi Rating	71
4.3 Implementasi Antarmuka	73
4.3.1 Antarmuka Sistem.....	73
4.3.2 Antarmuka <i>Home</i>	74
4.3.3 Antarmuka Detail <i>Item</i>	76
4.3.4 Antarmuka Rating <i>User</i>	76
4.3.5 Antarmuka Rekomendasi.....	77

BAB V PENGUJIAN DAN ANALISIS

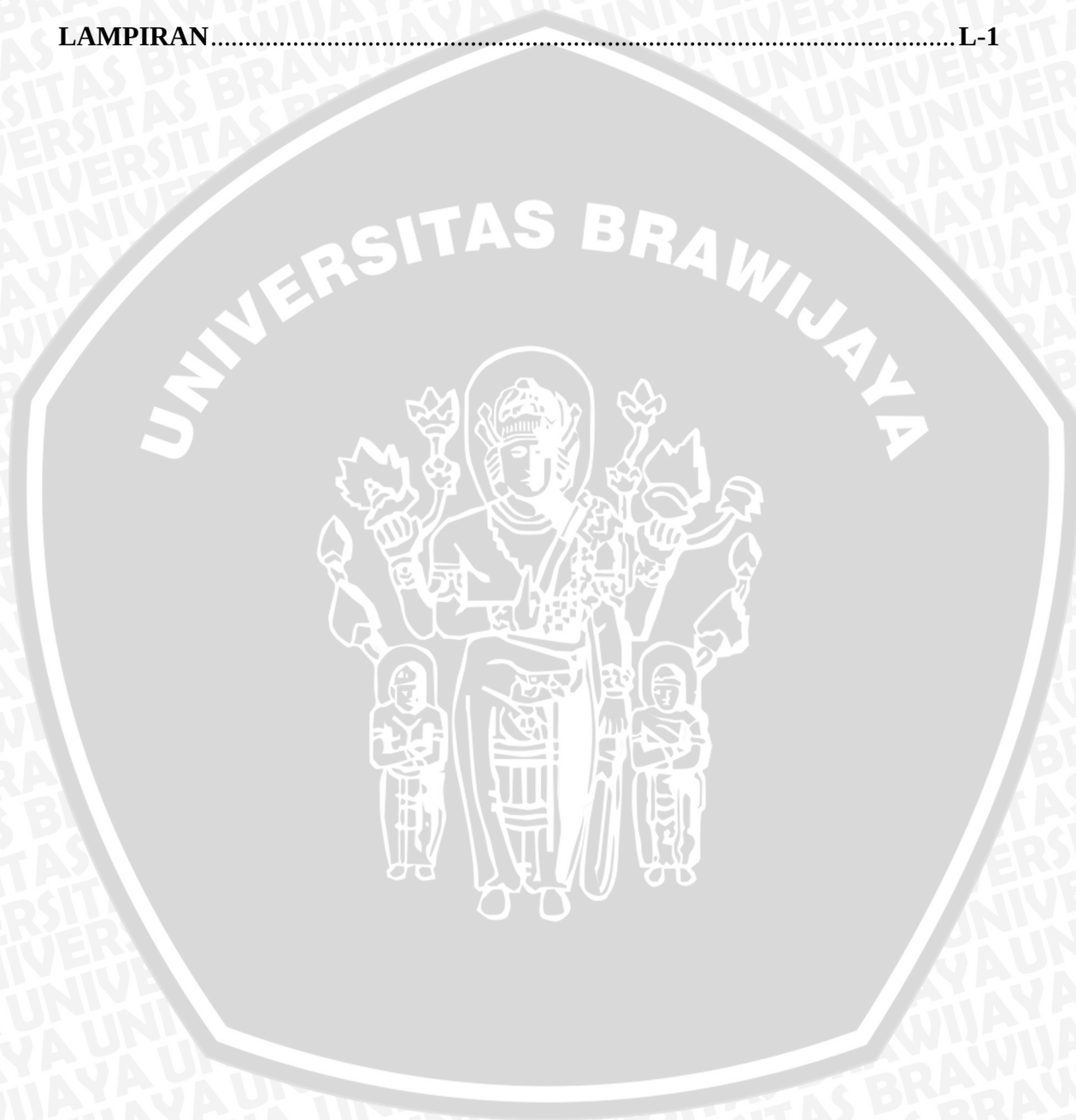
5.1 Pengujian.....	79
5.2 Hasil Uji	79
5.2.1 Hasil pengujian dan analisis pengaruh jumlah <i>cluster</i> terhadap tingkat akurasi hasil prediksi rating.....	79
5.2.2 Hasil pengujian dan analisis pengaruh koefisien <i>c</i> terhadap tingkat akurasi hasil prediksi rating.....	84
5.2.3 Hasil pengujian dan analisis pengaruh jumlah data latih dan data uji terhadap tingkat akurasi hasil prediksi rating	86

BAB VI PENUTUP

6.1 Kesimpulan	89
6.2 Saran.....	89

DAFTAR PUSTAKA	DP-1
----------------------	------

LAMPIRAN.....	L-1
---------------	-----



DAFTAR GAMBAR

	Halaman
Gambar 2.1 Arsitektur Sistem Rekomendasi	6
Gambar 2.2 <i>Linier Combination</i>	8
Gambar 2.3 <i>Sequential Combination</i>	9
Gambar 2.4 <i>Item-based Clustering Hybrid Method</i>	9
Gambar 3.1 Diagram Perancangan Proses	19
Gambar 3.2 <i>Flowchart</i> Proses <i>Clustering</i>	23
Gambar 3.3 <i>Flowchart</i> Jarak Item Dengan Pusat Cluster	24
Gambar 2.4 <i>Flowchart</i> Hitung Pro	25
Gambar 3.5 <i>Flowchart</i> <i>Similarity Item-rating</i>	28
Gambar 3.6 <i>Flowchart</i> <i>Similarity Group-rating</i>	30
Gambar 3.7 <i>Flowchart</i> <i>Linier Combination of Similiraty</i>	32
Gambar 3.8 <i>Flowchart</i> Prediksi Pada Item yang Pernah di Rating	34
Gambar 3.9 <i>Flowchart</i> Prediksi Pada Item Yang Belum Pernah di Rating	36
Gambar 3.10 Antarmuka Sistem	38
Gambar 3.11 Antarmuka <i>Home</i>	39
Gambar 3.12 Antarmuka Detail <i>Item</i>	40
Gambar 3.13 Antarmuka Rating <i>User</i>	41
Gambar 3.14 Antarmuka Rekomendasi	42
Gambar 4.1 <i>Source Code</i> Fungsi Hitung <i>Euclidian Distance</i>	62
Gambar 4.2 <i>Source Code</i> Fungsi Hitung <i>BCV/WCV</i>	63
Gambar 4.3 <i>Source Code</i> Fungsi Hitung <i>Pro</i>	64
Gambar 4.4 <i>Source Code</i> Create Tabel <i>Clustering</i>	64

Gambar 4.5 <i>Source Code</i> Fungsi Simpan Pro	65
Gambar 4.6 <i>Source Code</i> Fungsi Hitung <i>Similarity Item-rating</i>	66
Gambar 4.7 <i>Source Code</i> Fungsi Hitung <i>Similarity Group-rating</i>	68
Gambar 4.8 <i>Source Code</i> Fungsi Hitung <i>Linier Combination of Similarity</i>	68
Gambar 4.9 <i>Source Code</i> Create Tabel <i>Similarity</i>	69
Gambar 4.10 <i>Source Code</i> Fungsi Simpan <i>Similarity</i>	69
Gambar 4.11 <i>Source Code</i> Fungsi Hitung Prediksi Rating	71
Gambar 4.12 <i>Source Code</i> Create Tabel <i>Prediction</i>	71
Gambar 4.13 <i>Source Code</i> Fungsi Simpan Rekomendasi	72
Gambar 4.14 <i>Source Code</i> Antarmuka Sistem	73
Gambar 4.15 <i>Source Code</i> Antarmuka <i>Home</i>	74
Gambar 4.16 <i>Source Code</i> Antarmuka <i>Detail</i>	75
Gambar 4.17 <i>Source Code</i> Antarmuka <i>Rating User</i>	76
Gambar 4.18 Antarmuka Rekomendasi	77
Gambar 5.1 Grafik Pengaruh Jumlah <i>Cluster</i> Terhadap Tingkat Akurasi Prediksi Rating dengan Data Latih dan Data Uji yang Sama Pada Setiap Jumlah <i>Cluster</i>	81
Gambar 5.2 Grafik Pengaruh Jumlah <i>Cluster</i> Terhadap Tingkat Akurasi Prediksi Rating dengan Data Latih dan Data Uji yang Berbeda Pada Setiap Jumlah <i>Cluster</i> ..	83
Gambar 5.3 Grafik Pengaruh Koefisien <i>c</i> Terhadap Tingkat Akurasi Prediksi Rating.....	85
Gambar 5.4 Grafik Pengaruh Jumlah Data Latih dan Data Uji Terhadap Tingkat Akurasi Prediksi Rating	88

DAFTAR TABEL

	Halaman
Tabel 3.1 Statistik Dataset <i>GroupLens</i>	16
Tabel 3.2 Contoh Data <i>Item</i> dan <i>User</i>	20
Tabel 3.3 Matriks <i>Item-rating</i>	20
Tabel 3.4 Contoh <i>Genre</i> Film	21
Tabel 3.5 Rata-rata Rating Pada <i>Genre</i>	21
Tabel 3.6 Nilai <i>Genre</i>	22
Tabel 3.7 Matriks <i>Group-rating</i>	26
Tabel 3.8 Hasil Rekomendasi	37
Tabel 3.9 Atribut Tabel Data	42
Tabel 3.10 Atribut Tabel <i>Item</i>	43
Tabel 3.11 Atribut Tabel <i>User</i>	44
Tabel 3.12 Pengujian Pengaruh Jumlah <i>Cluster</i> Terhadap <i>MAE</i>	45
Tabel 3.13 Pengujian Pengaruh Koefisien <i>c</i> Terhadap <i>MAE</i>	46
Tabel 3.14 Pengujian Pengaruh Jumlah Data Latih dan Data Uji Terhadap <i>MAE</i>	47
Tabel 3.15 Data Perhitungan Manual.....	47
Tabel 3.16 Matriks <i>Item-rating</i>	48
Tabel 3.17 <i>Movie Genre</i>	49
Tabel 3.18 Rata-rata Rating <i>Genre</i>	50
Tabel 3.19 Data <i>User Movie</i>	50
Tabel 3.20 Jarak Data Terhadap Pusat <i>Cluster</i>	51
Tabel 3.21 Hasil <i>Cluster</i>	52
Tabel 3.22 Matriks <i>Group-rating</i>	53

Tabel 3.23 Rata-rata Rating <i>Item</i>	54
Tabel 3.24 Hasil Perhitungan <i>Similarity Item-rating</i>	55
Tabel 3.25 Rata-rata <i>Group-rating</i>	56
Tabel 3.26 Hasil Perhitungan <i>Similarity Group-rating</i>	57
Tabel 3.27 Hasil Perhitungan <i>Linier Combination of Similarity</i>	59
Tabel 3.28 Rekomendasi <i>Item</i> Bagi User 94	61
Tabel 5.1 Statistik Data Latih dan Data Uji yang Sama Pada Setiap Jumlah <i>Cluster</i> Pada Pengujian Pengaruh Jumlah <i>Cluster</i> Terhadap <i>MAE</i>	80
Tabel 5.2 Hasil Pengujian Pengaruh Jumlah <i>Cluster</i> Terhadap <i>MAE</i> dengan Data Latih dan Data Uji yang Sama Pada Setiap Jumlah <i>Cluster</i>	80
Tabel 5.3 Statistik Data Latih dan Data Uji yang Berbeda Pada Setiap <i>Cluster</i> Pada Pengujian Pengaruh Jumlah <i>Cluster</i> Terhadap <i>MAE</i>	82
Tabel 5.4 Hasil Pengujian Pengaruh Jumlah <i>Cluster</i> Terhadap <i>MAE</i> dengan Data Latih dan Data Uji yang Berbeda Pada Setiap Jumlah <i>Cluster</i>	83
Tabel 5.5 Statistik Data Latih dan Data Uji Pada Pengujian Pengaruh Koefisien <i>c</i> Terhadap <i>MAE</i>	84
Tabel 5.6 Hasil Pengujian Pengaruh Koefisien <i>c</i> Terhadap <i>MAE</i> Hasil Pengujian Pengaruh Jumlah <i>Cluster</i>	85
Tabel 5.7 Statistik Dataset Pengujian Pengaruh Jumlah Data Latih Dan Data Uji Terhadap <i>MAE</i>	86
Tabel 5.8 Hasil Pengujian Pengaruh Jumlah Data Latih Dan Data Uji Terhadap <i>MAE</i>	87

DAFTAR LAMPIRAN

- Lampiran 1. Hasil Pengujian Pengaruh Jumlah *Cluster* Terhadap Tingkat Akurasi Hasil Prediksi Rating
- Lampiran 2. Hasil Pengujian Pengaruh Koefisien c Terhadap Tingkat Akurasi Hasil Prediksi Rating
- Lampiran 3. Hasil Pengujian Pengaruh Jumlah Data Latih dan Data Uji Terhadap Tingkat Akurasi Hasil Prediksi Rating

